Project Report: Team Golden Sales

Introduction

Our current project is based on the sales and finances of an Electric Car making company named "Tesla". Tesla is an Electric Vehicle and clean energy supplier company, that is based in the United States. It manufactures cars, batteries, solar panels, etc., and is emerging to become one of the most valuable countries in the world. Founded in 2003, it produced its first car in the year 2009, and was IPO listed in 2010. It launched itself in the Electric Car domain in the year 2012, and has dominated the electric car space with its modern EV and driver assistant technology. Currently, its presence is seen in most of the developed first world countries, including USA, Canada, China and Europe.

As employees under the CFO of Tesla, we intend on advising the management regarding its expenditures and future sales. From its historic sales, budget and accounting information, we attempt to create a time series model to predict the sales of Tesla in various continents. Additionally, we attempted a production model to cater to the production capacity of Tesla and another model to predict the required Capital Expenditures to meet the sales, keeping a certain goal in mind. Climate change information, which was taken from NASA database, has also been added to the data to help advice the marketing team for associating climate change and the need for renewable energy to meet its marketing goals.

Data extraction

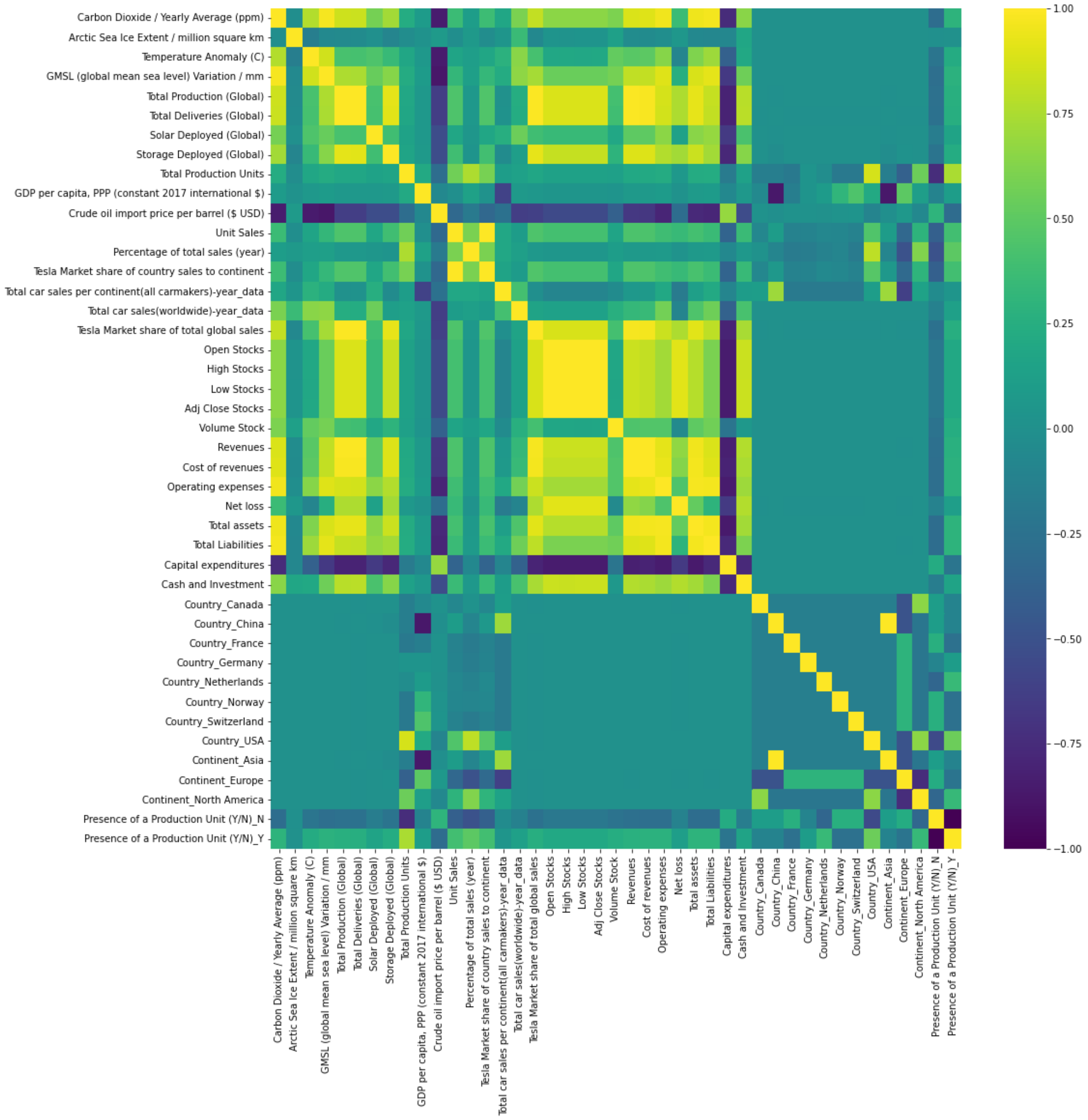
Our data contains an array of information that we felt were useful for carrying prediction models for Tesla. Since Tesla is a publicly listed company, the accounts information, including the sales information, were obtained from its website, starting from a period of 2011. These sales have been segregated country wise. Certain worldwide car sales information was also taken from data websites to understand the growth in the worldwide market share of Tesla. The factory information was taken from Wikipedia, and the share price information, along with its increasing volume sold were taken from an API named yfinance. Additionally, public data websites helped in providing us with the climate change information, GDP of countries and crude Oil prices. Certain type of information that we felt was important, like the Public belief in climate change, policies linked to EV in countries, import duties on cars within countries, standard fuel prices country wise, etc. could, unfortunately, not be found on the internet.

Data Cleaning

Our data contains certain missing information. Our first step, therefore, is to remove the columns that have more than 20% null values. Thus, we ended up removing columns like Fuel Prices, Public Belief in climate, etc. Our next step would be to replace the null values in the remaining columns with their most suitable values. For this purpose we have used the iterative imputer method provided in the sklearn library of Python. This Iterative Imputer imputes all the values corresponding to the values present in the datasets by iterating through all the rows and columns. At the end, we obtained non-null values for all our features, and were satisfied with proceeding forward.

Exploratory Data Analysis

In this section, we are going to view our dataset and understand the meaning behind our values. The analysis has also been conducted in the Power BI dashboard attached with the file. We will use the library Seaborn in Python for visualizing our data.



Observations: EDA

From the correlation matrix above we can see that:

1. Unit Sales are highly correlated with Cash and Investment, Total Liabilities and Assets, Net Loss, Operating Expenses, Revenues, Total Production and Stocks factors. This means that Unit Sales increase impacts the increase in other Tesla financial components. It also shows good correlation with the countries.
2. Most of the revenues and financial information look correlated with the investments Tesla has made over time, for example in Production Units, Solar deployed, etc. One can be a causation of the other.
3. Capital expenditure values (which are negative in our data set) also shows high correlation to most of the factors.
4. GDP and Crude Oil prices do not show very well correlation with the Unit Sales of Tesla. However, it shows correlation with the cars sold worldwide from all automakers.

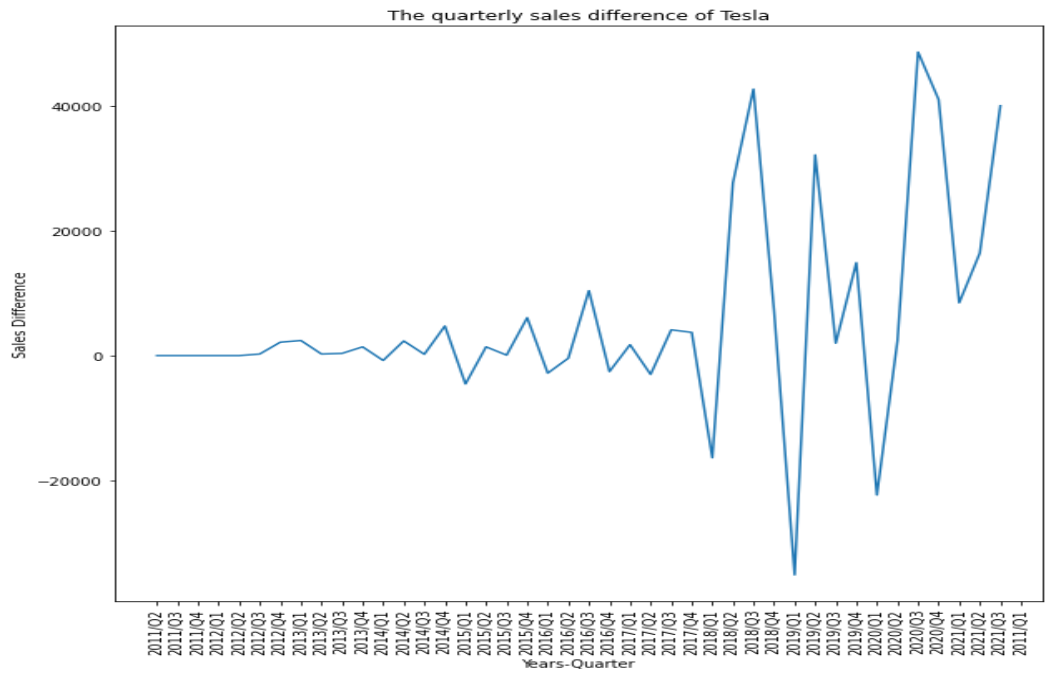
5. Climate change information does not show very significant correlation with the sales of Tesla. However, the share prices of Tesla are very much dependent upon these factors, signifying a good approach towards Tesla's marketing strategy of associating them with sales of Green goods.

Data Modelling

In this section, we will start carrying out modelling of the dataset. Three types of models will be placed in front of you, one involving the time series analysis and sales prediction modelling, the other associated with the production modelling, and the third that helps predict the capital expenditures for the company.

A) Worldwide Sales Prediction- Time Series Modelling

This Model is a time series model that was thought about while we wanted to undertake a sales prediction perspective. We used the Unit Sales of Tesla worldwide to understand the trends. For carrying out such modelling, we had to understand the stationarity of the sales of Tesla worldwide. Thus, upon having quarter wise sales information, we took out the difference in the sales of each quarter and plotted a graph which came out to be like the following:



As seen from above, a certain stationarity was present in the years around 2014 to 2017. Beyond that a similar trend was observed, but unfortunately, it shows bigger variance. It finally disrupts around 2020- 2021, where it is seen to not have any specific trend.

To carry out predictions for the worldwide Unit sales, we will be using Deep Learning Libraries like Keras for doing the predictions. Having obtained partial stationarity, we obtained lags within the sales over the time, and using 12 lags in total for our model, we got a f1 score of ~43%. This score is very small and cannot be relied upon for an optimal model. Thus, we had to drop our idea of using this model for our analysis.

B) Production Modelling

Our goal for this model is to obtain accurate predictions regarding the capacity of tesla for producing vehicles. This was based on our purpose for calculating the difference in the production output and the expected demand of Tesla in the market. For this modelling, we will use regression models to obtain the best fit for predicting this output. Feature selection will be done based on their correlations and p-value statistical tests, and accuracy for this model will be obtained.

Since production depends upon the factors internal to Tesla, we will only consider those features that are associated with Tesla. Thus, all accounting information, shares information, production information, and sales information was taken into perspective. Grouping of all these features was done based on their quarters, and packages like Pandas and Sklearn in Python were used for modelling our results.

After grouping, a min-max scaler was used on our datasets for scaling of our values, and then an ordinary regression was carried out using the stats library of Python. Upon obtaining its summary, we found out the variables that were affecting the regression curve through their P-Values. We considered a significance level of 10% for every variable, and then filtered out all those variables that had a p-value above ~0.10. This way we ended up with the stocks and revenue information of Tesla alone.

Upon test train splitting our dependent and independent variable, we used five different types of regression models which gave us the following results:

	Model	Scores
0	Lasso Regression	0.988084
1	Ridge Regression	0.975110
2	ElasticNetCV	0.975048
3	Random Forest Regressor	0.920311
4	Gradient Boosting Regressor	0.934824

From above, it is noticeable that we got models with a very high score. The Lasso Regression gave us a very good output of 0.98 and is tempting to consider. However, we were not satisfied with the features this model was based on. It was very highly dependent upon the Stock information which is, unfortunately, not under Tesla's control. So, overall, this model made little rational sense to us, and thus, we had to discard it.

C) Capital Expenditure Regression Model

In this model, we attempt to look into the factors affecting the Capital Expenditures. To meet our goal in advising the CFO on his expenditures considering a fixed percentage rise in sales, we build a model associating the two. In this model, we disregard all the climate, country, shares, and other information that is external to Tesla. All the features that will be selected in this model concern the internal functioning of the company.

We performed the same steps as we had done above for the production modelling. We again standardized our data, split it into test and train datasets, carried out feature selection by using 10% significance level, and modelled our regression curves. We ended up with using features like assets, liabilities, sales, net loss, and operating expense as the independent variables in our modelling. Ultimately, we got the following scores for our models:

	Model	Scores
0	Lasso Regression	0.918584
1	Ridge Regression	0.913942
2	ElasticNetCV	0.916252
3	Random Forest Regressor	0.583399
4	Gradient Boosting Regressor	0.590822

Here, we noticed that Lasso Regression gave us the best score of all the other regression models. Thus, we consider the Lasso regression in our analysis. The features selected for predicting the capital expenditures also make rational sense, and thus, we will finalize this model for our prediction analysis.

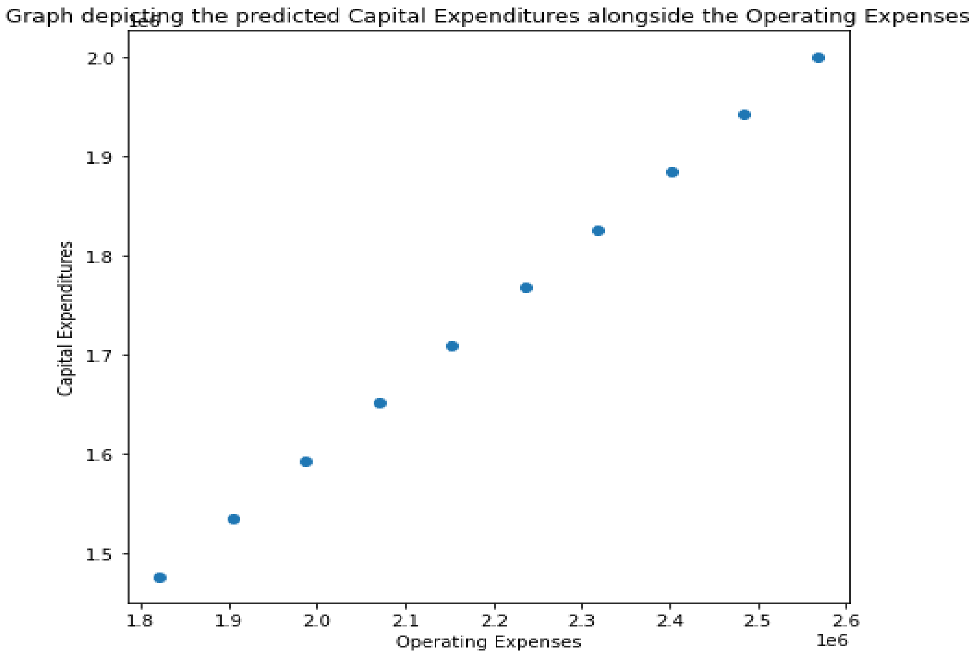
Results

Assumptions:

1. We have an ambitious project of increasing the sales by 5% of what it was last quarter. We significantly want to increase it by 2% after that.
2. We assume that the total liabilities and assets remain the same initially.
3. The operating expense we will increase by a few percent.
4. Assuming that the Net Loss is the same as for the last quarter.

Prediction A: Keeping a fixed sales goal,Net Loss,Liabilities and Assets as well as incresing Operatin g Expenses by 5 %, how the Capital Expenditure numbers are influenced?

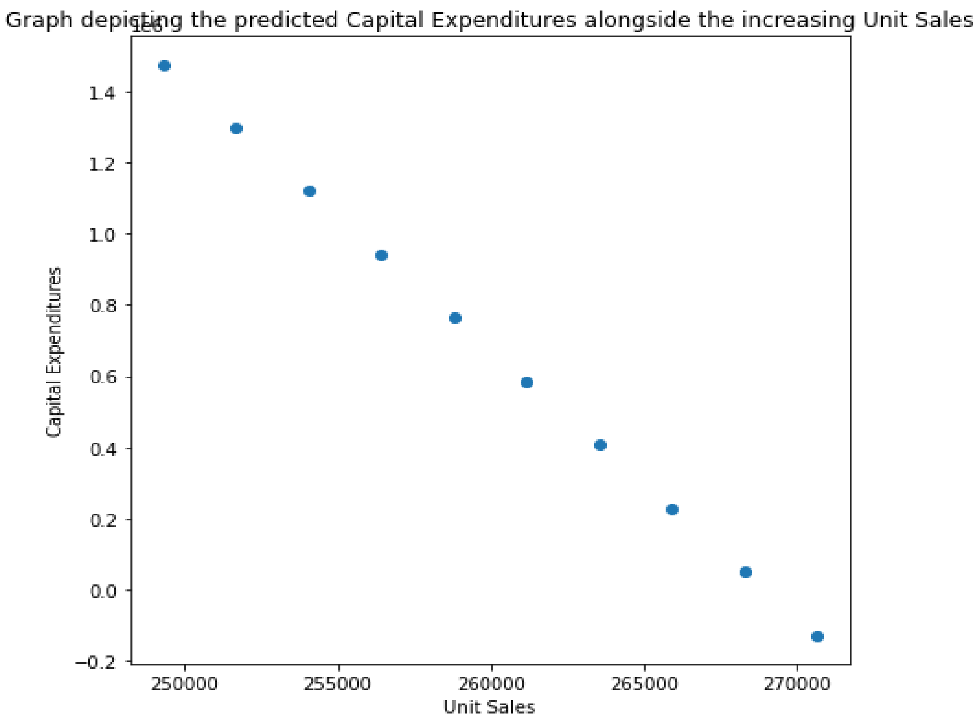
In this part of our results, we are considering keeping the financial factors like Assets, Net loss, etc. constant as the last quarter, sales as a 5% increase and then increasing the operating expenses subsequently to understand the relation. We obtained the following graphs:



From above we see that the Capital Expenditure has a steady linear rise, as we continue to increase our Operating Expenses to meet the production and sale demand.

Prediction B: Keeping a fixed Liabilities and Assets, Net Loss values and increasing Sales by 1% as well as Operating Expenses by 2%, how Capital Expenditure is influenced?

Here, we directly look at how the Capital Expenditures are related to the Unit Sales. We kept every factor constant and kept increasing the Unit Sales, first by 5%, and then subsequently by 1%. We obtained the following result:



As noticed, capital expenditures decrease with increasing Unit Sales.

Conclusion

Looking at the graph above we observe that Tesla's Capital Expenditure is decreasing while Unit sale s are increasing by 1%. At a sale margin of almost 13% increase from the last quarter value, keeping t he other factors constant, the value of the capital expenditures goes down to negative. A plausible e xplanation of this comes from the fact that as sales increase, the revenue generated increases too as a result, which would decrease the capital expenditure (coming out of Tesla's pockets). Eventually, it goes under good profit under ideal market conditions. Thus, our advice to the CFO would be to aim at a 13% increase in his/her company Electric Vehicle sales for better profit margins.

As seen from the correlation analysis, using the climate change statistics for marketing Tesla Electric Vehicles is a good approach as share prices of Tesla seem highly correlated to this aspect.

Acknowledgement

We would like to thank the University of Zurich and our faculty from Deloitte for giving us the opportunity for working on this project. We would especially like to thank our mentor Vladimir Popov for being patient and helpful to us throughout and making us reach our goal.

References and Bibliography

1. https://www.statista.com/topics/2086/tesla/#dossierKeyfigures
2. https://carsalesbase.com/volkswagen-group-us/
3. https://ir.tesla.com/#tab-quarterly-disclosure
4. https://climatecommunication.yale.edu/visualizations-data/ycom-us/
5. https://www.kaggle.com/mruanova/us-gasoline-and-diesel-retail-prices-19952021
6. https://carsalesbase.com/europe-tesla/
7. https://towardsdatascience.com/predicting-sales-611cb5a252de
8. https://en.wikipedia.org/wiki/List_of_Tesla_factories
9. https://ourworldindata.org/grapher/gdp-per-capita-worldbank?country=~CHE
10. https://data.oecd.org/energy/crude-oil-import-prices.htm